Curriculum vitæ: Felix Dangel ▼♀♀

I am a Postdoc at the Vector Institute in Toronto. My goal is to **build practical second-order optimizers to** accelerate the training of neural nets for language modelling and scientific ML, hand in hand with advancing automatic differentiation techniques and integrating them into next-generation ML libraries. More broadly, these techniques transfer to many other tasks built on second-order Taylor expansion (model merging and compression, uncertainty quantification, training data attribution, bi-level optimization, ...). During my PhD, I extended gradient backpropagation to efficiently extract higher-order information about the loss landscape of neural nets to improve their training. Before, I studied Physics at the University of Stuttgart with a main interest in simulating quantum many-body systems with tensor networks. I am passionate about

- developing automatic differentiation tricks to tackle efficient extraction of richer deep learning quantities, like second-order and per-sample information, and integrating that functionality into machine learning libraries,
- using these quantities to build better algorithms-specifically for, but not limited to, optimization-, and
- releasing code that empowers the community (for example cockpit, backpack, vivit, curvlinops, singd, einconv, sirfshampoo, unfoldNd).

Education

Start	Postdoctoral researcher , Vector Institute, Toronto
2025	With: Prof. Dr. Roger Grosse
2025	Postdoctoral researcher , Vector Institute, Toronto
_	Research statement: <i>Deep Learning Needs More Than Just the Gradient</i>
2023	With: Prof. Dr. Yaoliang Yu
2023 _ 2018	PhD in Computer Science , Max Planck Institute for Intelligent Systems & University of Tübingen Thesis: <i>Backpropagation beyond the Gradient</i> Advisor: Prof. Dr. Philipp Hennig
2018	Researcher , University of Stuttgart
_	Paper: <i>Topological invariants in dissipative extensions of the Su-Schrieffer-Heeger model</i>
2017	Host: Institute for Theoretical Physics 1
2017	MSc in Physics , University of Stuttgart
_	Thesis: <i>Bosonic many-body systems with topologically nontrivial phases subject to gain and loss</i>
2015	Advisor: P. D. Holger Cartarius
2015	BSc in Physics , University of Stutgart
_	Thesis: <i>Microscopic description of a coupling process for PT-symmetric Bose-Einstein condensates</i>
2012	Advisor: Prof. Dr. Günter Wunner

Publications

- [pre-print 2025] Improving Energy Natural Gradient Descent through Woodbury, Momentum, and Randomization
 - A. Guzmán-Cordero, F. Dangel, G. Goldshlager, M. Zeinhofer (arXiv)
- [pre-print 2025] *Collapsing Taylor Mode Automatic Differentiation* F. Dangel*, T. Siebert*, M. Zeinhofer, A. Walther (arXiv)
- [pre-print 2025] *Position: Curvature Matrices Should Be Democratized via Linear Operators* F. Dangel*, R. Eschenhagen*, W. Ormaniec, A. Fernandez, L. Tatzel, A. Kristiadi (arXiv | code)
- [pre-print 2025] *Spectral-factorized Positive-definite Curvature Learning for NN Training* W. Lin, F. Dangel, R. Eschenhagen, J. Bae, R. E. Turner, R. B. Grosse (arXiv)
- [ICML 2025 spotlight]: Fishers for Free? Approximating the Fisher Information Matrix by Recycling the Squared Gradient Accumulator Y. X. Li, F. Dangel, D. Tam, C. Raffel (pdf)
- [ICML 2025 spotlight] Hide & Seek: Transformer Symmetries Obscure Sharpness & Riemannian Geometry Finds It
 M. F. da Silva, F. Dangel, S. Oore (pdf, arXiv)
- [ICLR 2025 spotlight] What Does It Mean to Be a Transformer? Insights from a Theoretical Hessian Analysis
 W. Ormaniec, F. Dangel, S. P. Singh (pdf | arXiv | video)
- [NeurIPS 2024] *Kronecker-Factored Approximate Curvature for Physics-Informed Neural Networks* F. Dangel*, J. Mueller*, M. Zeinhofer* (pdf | arXiv | video | poster | code | slides)
- [NeurIPS 2024] Convolutions and More as Einsum: A Tensor Network Perspective with Advances for Second-Order Methods
 F. Dangel (pdf | arXiv | code | video | poster | slides)
- **[ICML 2024 workshop]** *Lowering PyTorch's Memory Consumption for Selective Differentiation* S. Bhatia, **F. Dangel** (pdf | arXiv | code | poster | bug report)
- [ICML 2024] Revisiting Scalable Hessian Diagonal Approximations for Applications in Reinforcement Learning
 M. Elsayed, H. Farrahi, F. Dangel, R. Mahmood (pdf | arXiv | code)
- **[ICML 2024]** Can We Remove the Square-Root in Adaptive Gradient Methods? A Second-Order Perspective W. Lin, F. Dangel, R. Eschenhagen, J. Bae, R. Turner, A. Makhzani (pdf | arXiv | poster | code)
- **[ICML 2024]** *Structured Inverse-Free Natural Gradient: Memory-Efficient & Numerically-Stable KFAC* W. Lin*, **F. Dangel***, R. Eschenhagen, K. Neklyudov, A. Kristiadi, R. Turner, A. Makhzani (pdf | arXiv | code | poster)
- [pre-print 2023] *On the Disconnect Between Theory and Practice of Overparametrized Neural Networks* J. Wenger, F. Dangel, A. Kristiadi (pdf | arXiv)
- [NeurIPS 2023] The Geometry of Neural Nets' Parameter Spaces Under Reparametrization A. Kristiadi, F. Dangel, P. Hennig (pdf | arXiv)
- [PhD thesis 2023] *Backpropagation Beyond the Gradient* F. Dangel (pdf | source | template)

- [TMLR 2022] ViViT: Curvature access through the generalized Gauss-Newton's low-rank structure F. Dangel*, L. Tatzel*, P. Hennig (pdf | journal | arXiv | code | www)
- [NeurIPS 2021] Cockpit: A Practical Debugging Tool for Training Deep Neural Networks F. Schneider*, F. Dangel*, P. Hennig (pdf | conference | arXiv | code | www | video)
- [ICLR 2020 spotlight] *BackPACK: Packing more into backprop* F. Dangel*, F. Kunstner*, P. Hennig (pdf | conference | arXiv | code | www | video)
- [AISTATS 2020] *Modular Block-diagonal Curvature Approximations for Feedforward Architectures* F. Dangel, S. Harmeling, P. Hennig (pdf | conference | arXiv | code | video)
- [Phys. Rev. A 2018] *Topological invariants in dissipative extensions of the Su-Schrieffer-Heeger model* F. Dangel*, M. Wagner*, H. Cartarius, J. Main, G. Wunner (pdf | journal | arXiv)
- [Acta Polytechnica 2018] *Numerical calculation of the complex Berry phase in non-Hermitian systems* M. Wagner*, F. Dangel*, H. Cartarius, J. Main, G. Wunner (pdf | journal | arXiv)
- [Master thesis 2017] Bosonic many-body systems with topologically nontrivial phases subject to gain and loss
 F. Dangel (pdf)
- [Bachelor thesis 2015] Mikroskopische Beschreibung eines Einkoppelprozesses fur PT-symmetrische Bose-Einstein-Kondensate
 F. Dangel (pdf, German only)

Talks & Workshops

- **[2025]** I gave an invited talk about accelerating Taylor mode automatic differentiation at the workshop 'Overparametrization, Regularization, Identifiability and Uncertainty in Machine Learning' hosted at the Mathematisches Forschungsinstitut Oberwolfach (slides)
- [2024] I presented my NeurIPS paper "Convolutions and More as Einsum" at the Vector Research Day in November (slides)
- **[2024]** I was given the opportunity to prepare a short presentation for Geoffrey Hinton about my joint work with Marvin F. da Silva and Sageev Oore for a Swedish TV event in October (footage: 1, 2)
- **[2024]** Invited talk at Cerebras Systems seminar (June 2024) and Graham Taylor's group meeting (July 2024) on "Convolutions Through The Lens of Tensor Networks" (slides)
- [2024] I gave a tutorial on "Large-scale Linear Algebra with Curvature Matrices" in Collin Raffel's group meeting in April (notes)
- [NeurIPS 2023] Workshop poster presentation at NeurIPS OPT23 on "Structured Inverse-Free Natural Gradient Memory-Efficient & Numerically-Stable KFAC for Large Neural Nets" (poster)
- [2023] Invited talk at Perimeter Institute Machine Learning Initiative seminar (December 2023) titled "Deep Learning Convolutions Through the Lens of Tensor Networks" (recording, slides)
- [2022] Poster presentation at the ELLIS Doctoral Symposium (EDS) 2022 in Alicante (poster)
- [2022] Invited talk at ELISE Theory Workshop on ML Fundamentals 2022 at EURECOM in Sofia Antipolis
- [2022] Poster presentation at the ELLIS Theory Workshop 2022 in Arenzano

- [2022] Session chair at the Franco-German Research and Innovation Network on AI, June 2022
- [2021] Co-organization of the ELLIS Doctoral Symposium (EDS) 2021 in Tübingen, held 2022 in Alicante
- [2019] Invited DL overview talk, seminar for Integrated Engineering students, DHBW CAS in Heilbronn
- [2017] Talk at the DPG Spring Meeting of the atomic Physics and quantum optics section 2017 in Mainz
- [2015] Participation at the "Ferienakademie 2015" summer school, organized by the TU Munich, the University of Stuttgart, and the FAU Erlangen in Sarntal (Northern Italy), talk about Lattice Boltzmann Methods.
- [2014] Participation at the "Ferienakademie 2014" summer school, organized by the TU Munich, the University of Stuttgart, and the FAU Erlangen in Sarntal (Northern Italy), talk about NMR & MRI

Teaching, Mentoring, Reviewing & Community Service

- **[2018-2022]** I taught seven (7) iterations of software development practicals. In these courses, three PhD students supervise ~15 students whose task is to develop a machine learning prediction system for the German soccer league over the course of one term (example). The overall workload for a student is ~180 hours and the focus lies heavily on teaching good software development practices.
- I have worked with various students:
 - [2024, ongoing] Disen Liao & Yihan Wang (undergraduate students at University of Waterloo, in cosupervision with Yaoliang Yu) on bi-level optimization problems data poisoning attacks.
 - [2024, ongoing] Andrés Guzmán-Cordero (Master student at University of Amsterdam, in co-supervision with Marius Zeinhofer and Gil Goldshlager) on using randomized linear algebra and Woodbury's formula to accelerate second-order optimizers for Physics-informed neural networks.
 - [2024-2025] Yu Xin Li (undergraduate student at University of Toronto, in co-supervision with Derek Tam and Colin Raffel) on applying gradient accumulators as Fisher proxies (ICML spotlight).
 - [2024-2025] Weronika Ormaniec (master student at ETH Zürich, in co-supervision with Sidak Pal Singh) worked on gaining insights into the loss landscape of transformers by theoretically investigating their Hessian (ICLR spotlight).
 - [2023-2024] Samarth Bhatia (master student) worked on randomized autodiff for convolutions and wrote an ICML workshop paper identifying a sub-optimality in PyTorch's automatic differentiation.
 - [2022] Elisabeth Knigge (high school student, summer internship) worked on making deep learning optimization methods more approachable to non-experts through visualization.
 - [2021-2022] Jessica Bader (research assistant) worked on broadening BackPACK's support for Kroneckerfactorized curvature for Bayesian deep learning. She wrote the interface for negative log-likelihood losses to support KFAC and to enable applications with their Laplace approximation via the laplacetorch library.
 - [2021] Tim Schäfer (Master thesis), added support for ResNets and recurrent architectures to Back-PACK. The underlying converter that makes these architectures compatible can easily be enabled through an optional argument while extending the model.
 - [2020-2021] Shrisha Bharadwaj (research assistant) improved BackPACK's code quality through additional tests, docstrings, and extended support for two-dimensional convolution to 1d and 3d.
 - [2019-2020] Paul Fischer (research project), implemented and analyzed Hessian backpropagation for batch normalization to speed up its Hessian-vector product, that can be slow (page 7), through structural knowledge.

- [2019] Christian Meier (Bachelor thesis): Activity prediction in smart home environments via Markov models.
- Served as area chair for International Conference for Machine Learning (ICML) (2025)
- I have reviewed for top-tier machine learning conferences and journals
 - Advances in Neural Information Processing (NeurIPS) (2020, 2021, 2022 (HITY workshop), 2023, 2024, 2024 (OPT workshop))
 - International Conference for Machine Learning (ICML) (2020, 2021, 2022, 2024)
 - International Conference on Learning Representations (ICLR) (2024)
 - Journal of Machine Learning Research (JMLR) (2021)
 - International Conference on Artificial Intelligence and Statistics (AISTATS) (2024)
- Served as reviewer for the Vector Scholarship in Artificial Intelligence 2023-2024
- [NeurIPS 2023] Represented Vector Institute as mentor at the Black in AI workshop's Q&A session
- [NeurIPS 2024] Served as mentor for the Black in AI workshop